```
   1 CGCCTCCCAG CGACTCTCGG CAGTGCCGGA GTCGGGTGGG TTGGCGGCTA
  51 TAAAGCTGGT AGCGAAGGGG AGGCGCCGCG GACTGTCCTT TCGTGGCTCA
 101 CTCCCTTTCC TCTGCTGCCG CTCGGTCACG CTTGCTCTTT CACCATGCCT
 151 GGATCACTTC CTTTGAATGC AGAAGCTTGC TGGCCAAAAG ATGTGGGAAT
 201 TGTTGCCCTT GAGATCTATT TTCCTTCTCA ATATGTTGAT CAAGCAGAGT
 251 TGGAAAAATA TGATGGTGTA GATGCTGGAA AGTATACCAT TGGCTTGGGC
 301 CAGGCCAAGA TGGGCTTCTG CACAGATAGA GAAGATATTA ACTCTCTTTG
 351 CATGACTGTG GTTCAGAATC TTATGGAGAG AAATAACCTT TCCTATGATT
 401 GCATTGGGCG GCTGGAAGTT GGAACAGAGA CAATCATCGA CAAATCAAAG
 451 TCTGTGAAGA CTAATTTGAT GCAGCTGTTT GAAGAGTCTG GGAATACAGA
 501 TATAGAAGGA ATCGACACAA CTAATGCATG CTATGGAGGC ACAGCTGCTG
 551 TCTTCAATGC TGTTAACTGG ATTGAGTCCA GCTCTTGGGA TGGGCTTCGT
 601 GGGACACATA TGCAACATGC CTATGATTTT TACAAGCCTG ATATGCTATC
 651 TGAATATCCT ATAGTAGATG GAAAACTCTC CATACAGTGC TACCTCAGTG
 701 CATTAGACCG CTGCTACTCT GTCTACTGCA AAAAGATCCA TGCCCAGTGG
 751 CAGAAAGAGG GAAATGATAA AGATTTTACC TTGAATGATT TTGGCTTCAT
 801 GATCTTTCAC TCACCATATT GTAAACTGGT TCAGAAATCT CTAGCTCGGA
 851 TGTTGCTGAA TGACTTCCTT AATGACCAGA ATAGAGATAA AAATAGTATC
 901 TATAGTGGCC TGGAAGCCTT TGGGGATGTT AAATTAGAAG ACACCTACTT
 951 TGATAGAGAT GTGGAGAAGG CATTTATGAA GGCTAGCTCT GAACTCTTCA
1001 GTCAGAAAAC AAAGGCATCT TTACTTGTAT CAAATCAAAA TGGAAATATG
1051 TACACATCTT CAGTATATGG TTCCCTTGCA TCTGTTCTAG CACAGTACTC
1101 ACCTCAGCAA TTAGCAGGGA AGAGAATTGG AGTGTTTTCT TATGGTTCTG
1151 GTTTGGCTGC CACTCTGTAC TCTCTTAAAG TCACACAAGA TGCTACACCG
1201 GGGTCTGCTC TTGATAAAAT AACAGCAAGT TTATGTGATC TTAAATCAAG
1251 GCTTGATTCA AGAACTGGTG TGGCACCAGA TGTCTTCGCT GAAAACATGA
1301 AGCTCAGAGA GGACACCCAT CATTTGGTCA ACTATATTCC CCAGGGTTCA
1351 ATAGATTCAC TCTTTGAAGG AACGTGGTAC TTAGTTAGGG TGGATGAAAA
1401 GCACAGAAGA ACTTACGCTC GGCGTCCCAC TCCAAATGAT GACACTTTGG
1451 ATGAAGGAGT AGGACTTGTG CATTCAAACA TAGCAACTGA GCATATTCCA
1501 AGCCCTGCCA AGAAAGTACC AAGACTCCCT GCCACAGCAG CAGAACCTGA
1551 AGCAGCTGTC ATTAGTAATG GGGAACATTA AGATACTCTG TGAGGTGCAA
1601 GACTTCAGGG TGGGGTGGGC ATGGGGTGGG GGTATGGGAA CAGTTGGAGG
1651 AATGGGATAT CTGGGGATAA TTTTAAAGGA TTACATGTTA TGTAAATTTT
1701 TATGTGACTG ACATGGAGCC TGGATGACTA TCGTGTACTT GGGAAAGTCT
1751 CTTTGCTCTA TTTGCTGACA TGCTTCCTGT TGTGGTCTGG CCAATGCCAA
1801 ATGTACTCGA ATGATGTTAA GGGCTCTGTA AAACTTCATA CCTCTTTGGC
1851 CATTTGTATG CATGATGTTT GGTTTTTAAA CATGGTATAA TGAATTGTGT
1901 ACTTCTGTCA GAAGAAAGCA GAGGTACTAA TCTCCAATTA AAAAATTTTT
1951 TAACATGTAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2001 AA  (SEQ ID NO:1)
```

FIGURE 1A

FEATURES:

5'UTR: 1-144

Start Codon: 145

Stop Codon: 1579

3'UTR: 1582

Homologous proteins:

Top 10 BLAST Hits

	Score	E
CRA|18000004923628 /altid=gi|4504429 /def=ref|NP_002121.1| 3-hy...	961	0.0
CRA|18000004928954 /altid=gi|284048 /def=pir||S27197 hydroxymet...	945	0.0
CRA|18000004939530 /altid=gi|8393538 /def=ref|NP_058964.1| 3-hy...	915	0.0
CRA|18000004933126 /altid=gi|123332 /def=sp|P13704|HMCS_CRIGR H...	912	0.0
CRA|18000004944250 /altid=gi|123331 /def=sp|P23228|HMCS_CHICK H...	811	0.0
CRA|18000004996464 /altid=gi|86312 /def=pir||S13887 hydroxymeth...	810	0.0
CRA|108000024648192 /altid=gi|12731376 /def=ref|XP_011295.1| 3-...	673	0.0
CRA|18000004879762 /altid=gi|1708233 /def=sp|P54870|HMC2_BLAGE ...	489	e-137
CRA|18000005054533 /altid=gi|7436678 /def=pir||T09688 hydroxyme...	384	e-105
CRA|271273992 /altid=gi|7799986 /def=gb|AAF69804.1|AF148847_1 (...	377	e-103

BLAST dbEST hits:

	Score	E
gi|10952250 /dataset=dbest /taxon=96...	1247	0.0
gi|6854981 /dataset=dbest /taxon=9606...	1068	0.0
gi|10992587 /dataset=dbest /taxon=96...	894	0.0
gi|12762375 /dataset=dbest /taxon=960...	890	0.0
gi|11125858 /dataset=dbest /taxon=96...	890	0.0
gi|7376370 /dataset=dbest /taxon=9606...	890	0.0
gi|10991736 /dataset=dbest /taxon=96...	884	0.0
gi|10992783 /dataset=dbest /taxon=96...	882	0.0
gi|10990968 /dataset=dbest /taxon=96...	882	0.0

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

From BLAST dbEST hits:

gi|10952250 Teratocarcinoma

gi|6854981 Fetal brain

gi|10992587 Teratocarcinoma

gi|12762375 Liver- adenocarcinoma

gi|11125858 Lung small cell carcinoma

gi|7376370 Genitourinary tract

FIGURE 1B

gi|10991736 Teratocarcinoma

gi|10992783 Teratocarcinoma

gi|10990968 Teratocarcinoma neuronal repcursor cells

From tissue screening panels:

Whole liver

FIGURE 1C

```
  1 MPGSLPLNAE ACWPKDVGIV ALEIYFPSQY VDQAELEKYD GVDAGKYTIG
 51 LGQAKMGFCT DREDINSLCM TVVQNLMERN NLSYDCIGRL EVGTETIIDK
101 SKSVKINLMQ LFEESGNIDI EGIDTINACY GGTAAVFNAV NWIESSSWDG
151 LRGIHMQHAY DFYKPDMLSE YPIVDGKLSI QCYLSALDRC YSVYCKKIHA
201 QWQKEGNDKD FTLNDFGFMI PHSPYCKLVQ KSLARMLLND FLNDQNRDKN
251 SIYSGLEAFG DVKLEDTYFD RDVEKAFMKA SSELFSQKTK ASLLVSNQNG
301 NMYISSVYGS LASVLAQYSP QQLAGKRIGV FSYGSGLAAT LYSLKVTQDA
351 TPGSALDKIT ASLCDLKSRL DSRTGVAPDV FAENMKLRED THHLVNYIPQ
401 GSIDSLFEGT WYLVRVDEKH RRTYARRPTP NDDTLDEGVG LVHSNIATEH
451 IPSPAKKVPR LPATAAEPEA AVISNGEH  (SEQ ID NO:2)
```

FEATURES:
Functional domains and key regions:
[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site

81-84 NLSY

[2] PDOC00004 PS00004 CAMP_PHOSPHO_SITE
cAMP- and cGMP-dependent protein kinase phosphorylation site

426-429 RRPT

[3] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site

Number of matches: 4

1	60-62	TDR
2	103-105	SVK
3	286-288	SQK
4	343-345	SLK

[4] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

Number of matches: 16

1	60-63	TDRE
2	96-99	TIID
3	118-121	TDIE
4	146-149	SSWD
5	185-188	SALD
6	354-357	SALD
7	212-215	TLND
8	254-257	SGLE
9	267-270	TYFD
10	185-188	SALD

FIGURE 2A

11	354-357	SALD
12	362-365	SLCD
13	368-371	SRLD
14	405-408	SLFE
15	429-432	TPND
16	434-437	TLDE

[5] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

Number of matches: 11

1	41-46	GVDAGK
2	50-55	GLGQAK
3	122-127	GIDTIN
4	131-136	GGTAAV
5	150-155	GLRGIH
6	300-305	GNMYIS
7	309-314	GSLASV
8	334-339	GSGLAA
9	336-341	GLAATL
10	401-406	GSIDSL
11	440-445	GLVHSN

[6] PDOC00009 PS00009 AMIDATION
Amidation site

324-327 AGKR

[7] PDOC00942 PS01226 HMG_COA_SYNTHASE
Hydroxymethylglutaryl-coenzyme A synthase active site

117-132 NIDIEGIDTINACYGG

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	300	320	0.990	Putative
2	327	347	1.033	Certain

FIGURE 2B

```
BLAST Alignment to Top Hit:
>CRA|18000004923628 /altid=gi|4504429 /def=ref|NP_002121.1|
          3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1
          (soluble); 3-hydroxy-3-methylglutaryl-Coenzyme A synthase
          1 [Homo sapiens] /org=Homo sapiens /taxon=9606
          /dataset=nraa /length=520
        Length = 520

 Score =  961 bits (2458), Expect = 0.0
 Identities = 478/520 (91%), Positives = 478/520 (91%), Gaps = 42/520 (8%)
 Frame = +1

Query: 145  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 324
            MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT
Sbjct: 1    MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60

Query: 325  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 504
            DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI
Sbjct: 61   DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Query: 505  EGIDTTNACYGGTAAVFNAVNWIESSSWDG---------------------------- 594
            EGIDTTNACYGGTAAVFNAVNWIESSSWDG
Sbjct: 121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYATGNARPTGGVGAVALL 180

Query: 595  ------------LRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 738
                        LRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI
Sbjct: 181  IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Query: 739  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 918
            HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA
Sbjct: 241  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 300

Query: 919  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 1098
            FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY
Sbjct: 301  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

Query: 1099 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 1278
            SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP
Sbjct: 361  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420

Query: 1279 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGIWYLVRVDEKHRRTYARRPTPNDDTLDEG 1458
            DVFAENMKLREDTHHLVNYIPQGSIDSLFEGIWYLVRVDEKHRRTYARRPTPNDDTLDEG
Sbjct: 421  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGIWYLVRVDEKHRRTYARRPTPNDDTLDEG 480
```

FIGURE 2C

```
Query: 1459 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 1578
           VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH
Sbjct: 481  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520  (SEQ ID NO:4)


>CRA|18000004928954 /altid=gi|284048 /def=pir||S27197
           hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5),
           cytosolic, fibroblast isoform - human /org=human
           /taxon=9606 /dataset=nraa /length=520
         Length = 520

 Score =  945 bits (2417), Expect = 0.0
 Identities = 471/518 (90%), Positives = 472/518 (90%), Gaps = 42/518 (8%)
 Frame = +1

Query: 145  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 324
           MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT
Sbjct: 1    MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60

Query: 325  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 504
           DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI
Sbjct: 61   DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Query: 505  EGIDTTNACYGGTAAVFNAVNWIESSSWDG------------------------------ 594
           EGIDTTNACYGGTAAVFNAVNWIESSSWDG
Sbjct: 121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYATGNARPTGGVGAVALL 180

Query: 595  ------------LRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 738
                        LRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI
Sbjct: 181  IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Query: 739  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 918
           HAQWQKE ND DFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGL+A
Sbjct: 241  HAQWQKEANDNDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLKA 300

Query: 919  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 1098
           FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY
Sbjct: 301  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

Query: 1099 SPQQLAGKRIGVFSYGSGLAATLYSLKVIQDATPGSALDKITASLCDLKSRLDSRTGVAP 1278
           SPQ LAGKRIGVFSYGSGLAATLYSLKVIQDATPGSALDKITASLCDLKSRLDSRTGVA
Sbjct: 361  SPQHLAGKRIGVFSYGSGLAATLYSLKVIQDATPGSALDKITASLCDLKSRLDSRTGVAQ 420
```

FIGURE 2D

```
Query: 1279 DVFAENMKLREDIHHLVNYIPQGSIDSLFEGIWYLVRVDEKHRRTYARRPTPNDDILDEG 1458
           DVFAENMKLREDIHHLVNYIPQGSIDSLFEGIWYLVRVDEKHRRTYARRPTPNDDILDEG
Sbjct: 421  DVFAENMKLREDIHHLVNYIPQGSIDSLFEGIWYLVRVDEKHRRTYARRPTPNDDILDEG 480

Query: 1459 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNG 1572
           VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNG
Sbjct: 481  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNG 518  (SEQ ID NO:5)
```

Hmmer search results (Pfam):

Model	Description	Score	E-value	N
PF01154	Hydroxymethylglutaryl-coenzyme A synthase	1050.3	0	2

Parsed for domains:

Model	Domain	seq-f	seq-t		hmm-f	hmm-t		score	E-value
PF01154	1/2	13	150	..	1	138	[.	409.2	3.8e-119
PF01154	2/2	151	427	..	181	461	.]	635.2	3.6e-187

FIGURE 2E

```
   1 CCATTTTTCC CGCCATCACT GTCTTTAAAT TAGTCCATCG GAATTAGTTT
  51 AGCCTGTGCA GTCTAACCCT AGCCAATAAG GGAACGACAC AGCAGTGGGG
 101 ACCACGTGCG TCAGGAATAA GAACCCCTTT CCCTCCCTCG TCCAAGTGTG
 151 CACTCACCAT TGCTCCATCT GTAAGGGTGC ACCCTTCTAT AGAAGTACCT
 201 TGCCTTGCTG AGAATTAAAA AGAAAATTTT ATATTCGACT GCTATTTCTT
 251 TTGCAGCATG GAAACTTTAT TTATAACAAG ATCTTCTGTA TCTAATTACT
 301 AACCCTTTTT GTTCTCCATT GCTTGGCTTC CCAGTAATCA ATAATCATGC
 351 TCACTTTGCT TAATTGAAGA TTAACGTGAT CAAAAAGACG GTCTGTTCCT
 401 TGTAGAAATT TCCGGTTGTG TAAGATGGTC ATTCTCATGA CCGTCTGGCT
 451 AATCATTTCC CATTATGTAC TCCTGGAGTT GGAATTATTT GCGATTCCTA
 501 ACGACAAAAC TGTATCTTCT TTCTTGTGTT TGTCCTTACT GCCTTTCAGC
 551 ATATTCCAAT ATGCCAAGAA TTTTAATCTC CTACCCCACC CCAAATTGCT
 601 GTTGATCATA ATCAGGCAAT GTCTCTCTCT CTGTTTACTA TCTAGTTACT
 651 TTACATACAT ATGAAGTGAG TCATGGGCAA TACTGTGGAA TGGAAATCAT
 701 TACTGAGTGG TCCTCTTCCC CCAAGTCATT TATGCCACCA CTTCACAGTG
 751 GTTCCATTTC CAATATATTT TGCCACTTTG CTGCTGAGAA TGTGTCTTAC
 801 TAGGTTAGCA TCTATAGTGG TTAAAAGAAT CTCCCATAAC AATAATTGTG
 851 TGAATCACAG AATTACCAAT GACCCCTTAT CAATAGCATT CCTGTTAATT
 901 AAATTGAGAT GGGGAGAGAT ACAAACAACT CCGAACCTCA CTCATGGTCC
 951 CCCACCAAAG CTAAGTATTA TGGCTTCTCT CTCTGACCAG ATAGAGGCAG
1001 AGTTTATTGC AAAGCCACAA GTGTCCTCCT TTGGATTCCC CCAAATAGTG
1051 TTTCAGTGAA TTCCTCTAGC TTGAATTGCT CCTCTCTATT TGCTGGGGGA
1101 GTTAGGCAGT CCGTATCCGA TGGATTTACT ATGCCGACAA TTACGTGGCC
1151 TTTCCACAGC CTTTTACTTG GCAGGTACCA CATATGAAGC TTAGAAGATA
1201 CAGTGGGCAA CAGGCCAAAT GGAGTCCCTT TCCTCAGAGT GCATGGCCTG
1251 GCAAAAATCC TTGAATTCAG TATCAACTTC CCTTCACAGG CAAGGCTCTG
1301 CACCCTCCCC ACGGATGCCT AATCCTGAAA CCATTTTGTT TTAGGTTTAG
1351 TTAGAAAGCT TTGTCTCAAG AGCACTTTTG TTTGTTCTGT TTTCTTTAAG
1401 TCAAGGTAGT TTTGAATAAA GGAGACAATN ATTTGAGTAT TTACAAATCG
1451 GGTATTTAGA CTATTTACAC ATATACAAGT TCTGGGTGAA GTATTCTGCT
1501 CCAATTTGCA ATCTACGCAC ACTTTGCTAG AAAACGTTAA GACTGAATTC
1551 AAATCAAGTA CAGTATTTCA GAAATCTTTC AGGTGAAGCC TAGTTCTGGT
1601 TGCTAGGCAA CCTGACAGAC TCCCAAGCTG GGACCACCTC GCCTCCCACA
1651 TTTGACCATC TCTCCAGCGG TGGGACGCGG AGTACCCATT GGCCCGCATC
1701 TCCTCTCACT TAGTCCCAAT TGGTCGGAGA ACCTCTCACT CCGCTCCCGT
1751 TGGCTCTCGC CGTATCTCGC AGCTCCGTCA TTGGCAACTG GGCTCTCGTG
1801 CCACCTCACG TCAGTCTCTC ACACCACTTC CTCGGCCCTG AGACTTTGTC
1851 CCCGCCTCTT CTCCCCGCCC TTCCAGCCAC GAGGGAAAAT CCTAGCGAGT
1901 CATCGCCTCT AGTTTCCTTT TGATTGGTAG AAGCCGGACT GGGGGCGCGG
1951 CGCTGCCGGG CAACTCTACC GGCCGCGATT GGCTGTGGGA GCCACCGTCC
2001 CGCCTCCCAG CGACTCTCGG CGGTGCCGGA GTCGGGTGGG TTGGCGGCTA
2051 TAAAGCTGGT GGCGAAGGGG AGGCGCCGCG GACTGTCCTT TCGTGGCTCA
2101 CTCCCTTTCC TCTGCTGCCG CTCGGTCACG CTTGGTGAGT GTCCCGCGCT
2151 GGGGAGTAGA ACTGGGCTGC GGAGGTGCCG CGGGCGGGGT GTGGGCCAGA
2201 CAGAGGCGGT GTCCTTGACT AGGCCCGAAG GAGCTGGGGC TCTGGGTCAG
2251 GACGTAGGCG TGGACTTTGC CCGGGAGGAT GGGGCACCGT GAGCGGGGCC
```

FIGURE 3A

2301 GGGCGGGGGT TCCCTCGTGA GGGACCTGAG GCCGACCGTA GCGGATCTGA
2351 GAAGATCCGA GAACACAGGC GAGTCGCGGA GGGGAGAACG CGAGAGGGCG
2401 TTGAGGTCTA GGTATTCTAA CGACAGAGGA GTTGGAGGTG CCAGAGAGGC
2451 AGCTGTGACC GCCTAGAGGT GAGTGGGGGG TGTCAGGAGG GGGAGAGAAG
2501 ACAGTTGGGC TACCAAGGCG TTTCCAGAGC GTTGGTTAAG GGTGGACGCC
2551 AAAGGATGGG CAAGATCCTC TTTAGACGGA GGCTGGTAGG TTCGCAGGGG
2601 GTGTGTCCTG CTGCCACATA TAGAGTTGAT GGAAAGAAGG GAAGTGGGTA
2651 GCATTACTTT TCTTCCTCAG CTCAGGTGCA AGAAAGCGTT CACAACCGTG
2701 ATTTAGACCT GGCTAAGTAC TGGGGCTCAG TCTGTACTTG CTTCAAATCT
2751 CATAGATCAC TGCCTCCCGC CTTCCTGCCT CCATATTTTT TTTTGTCTAC
2801 GTTTTAAAAA ATAGGCTTCC TTGGTGTTCT GAAATCCCAC ATCTCTCTCC
2851 TACTAATACC TTCGGGACCA GCTTTAGGTG ATACAGTGTA ATGGGCAGGC
2901 ACTCACAGAG TCCTCCCACA AATAGGTTTT GGATTAAGCT AAGGATATTT
2951 CAAAGCAAGT ATATGGAGTC TTTGAAAACC CACGTCTGGC CTTGACCAGT
3001 GGTAGAGAAA CGATTATTCT GATCCACTCT GGAGGAGGGA TTTGGGGAAC
3051 AAATAATGTG AGGTTGTGCC TGTTTGTCAT GCTTGTCCCT ATGGCCTTAG
3101 CCTTAAGGCA TCAGTAGCTG CTTTCACTGC TCACCTCTGC TGCAGCTCCC
3151 CACCTTCCCG AGGATGCTCT TGCCACCTGC TGCAGTAGGA TGATGTGTTC
3201 TGGTTGCTGC TAACTAACAT TTGCTCTGTT TTAGGCATGA ATATGAAAAA
3251 CAATGACAAG ATAAACAACA AAATTAAGAC AAATGGAAGT GCTCCTAGAG
3301 TTAACAGATT TTTCCTTCTG AGATGTGTTT TGGACTTTAT TGCACAGATA
3351 CTATTAGATG AGAGGCAGTT GAAAGTCGTT AACATTACCC GTGTCAGTAG
3401 TTCTTTGCAC TTGAGACACC TAAGCAGCTT GTGTTCTTTA AACTTTATTT
3451 TAAAATTGCA GTTATTTTTG TGTGAAGAAG GGGGCAGGGA TAGCATACCT
3501 TATGGGAAGA GAGAAAGGCT TTCTTTGTGT CTACCTTTGT AGATATTTCT
3551 CACCTAAGTT TGTAAGTTTG CCCTTTATTC GGTTCTACTT TAGTTCAGCT
3601 CAATTCTAGT ATAATCATCA GTAACCCCAG CACTCAGAAG GTCTGACTTA
3651 CGCTGTGGGG AGGGAGTGTA AAAGGATATT TTATGTTTGG AGCCATAGGC
3701 CACATCATTT GGGCCTTGTT TTAATTTTGT TTTTCATCTT AAATATCCCT
3751 CCAGATTGCT TTTACATCTT GTTTCTTTTA ACTGTGGATT GATTTTGAGA
3801 TTTTGACTTA GATTTTAGAT AGCTTTTCTC AGAAGAAATA AACGCAAAAA
3851 CCCGATATTG TTGTAACATC AGTTTCCTGT GTCCTCTAGA ATCATTTAAA
3901 ACCTGGTTGG ATCTTCCATA ATCCAGTGGA ATTGGATATG AGATGTAGCT
3951 GGAGAAGTTT GTTTTGCTAC ATATCAGAAT CTCCAATTAG TTTCATTTAG
4001 AAAGGAATAT AGCCTTATAA TTTTATGCTG GGTTACTGTG GAACCAAATA
4051 TCATAGAAGG ATGTGTGATA TTTTTATGTT TTTCAAGAAG GTAGTATAGA
4101 TTTAAAAGGT GGGATACATA TTACCTGTCC TAATGATAGG ACTAGATTTT
4151 TTTTTTTTTT TTTTTTGGGG AGACAGAATC TCGCTCTGTC GCCCAAGCTG
4201 GAGTGCAGCA GCGTGATCTC GGCTCACTGC AACTTATGCC TCCCAGTGAT
4251 TCTCCTGCCT CAGCCTCCCA AGTAGCTGGG ACTACCGGCA TGTGCCACCA
4301 CACCCAGCTA ATTTTTTTGT ATTTTTAGAA GAGATGGGGT GTCACCATGT
4351 TGGTCAGACT GGTCTTGAAC TCCTGACCTC AAATGATCCG TCCGCCTTGG
4401 CCTCCCAAAG TGCTGAGATT ACAGGCGTGA GCCACCATGC CTGGCTAGAA
4451 CTAGACTCTT AATCTCTTCA TCCTAATGCA TGGCGTGTGT TGATGTTCAC
4501 TTAATGTCTG TCAACTGGGT GTAGTTACAC CAGTAGCGGA GAGGCTAATC
4551 TTTGAAAGCC TGAAGTGTTG TCTTCATCTT TGCAGGGTTT TTAGTTGTGG

FIGURE 3B

4601 GTGCATATGG GAATGATTGT AAGACCAACA AATGTTTTCT GATTCCATAT
4651 GGGCTTCTTA CATTTTTCAC CTTGGAATCT GGGAACAATT GAAACCTACC
4701 ATATGCCTTG AACAGTAGCA GTAAAGAGCC AGTTTCTTTA AACTAGACAT
4751 TATGGTGCTG CAGCTCATCT CAAAACTGAT AGCAGGCTAC TCTGGACACA
4801 CTACATATAG AGTAGCCCTG CTCTGCAAGG AGCAGTAATA AATTAAAAAA
4851 AAAATTAAAA AGTGATAGCA GAAAGCACTT ACTACTGAGG GCTGCTACAA
4901 GTATTAAATC TAAAAGATTT GTCCTCTAGT AGTTATAACT CCAAATTCAG
4951 CCACTGAAAA ATGTGACATT TGAGTACCCT TTACTTCAAG GTCTCAAAGG
5001 GATTTCAAAA AATCAAAATA TATAGCCCCT CTCCCAAAAG AAGTGTAGGA
5051 ATCCTGTATG GATAAGAAGA CTGCCCATAA CTAGTTTTCC ATAGAGAGTA
5101 GGCTATGTAG ACTTGGGTAT GAATGACCTA CCTCTGTAGA AGTGCAGGTC
5151 CCTGATTAGA AAACTTATTT TCTGTGTGAT TTATCGAGGA AAGCTTCCAG
5201 GAAGAGGTGA CTTAGAACAG GGCCTTGAAG ATGAGTAGAA TCTCTGATAC
5251 GCAGACCAGT AACTCTGGGA GGAGGCAGGG ATGTCCATGC TTTTTACTTG
5301 GAGAACTATA CCAGAGTGTA CAGGTTTGAG CAAGTCTTTC TTAACATTAG
5351 TTTTTACTTG CTTGCTCCTA AGGAGGAAAG GTTGCCAACT TGTTCTTAAT
5401 TTCCTAGATT TATCTCCTGT AACAATGAGA AAGATCAATA GGTAACTGTT
5451 TATATTTTAT AGTTTACATA CCAAAATGTG TAGGCAATGA ACTTCTCCAA
5501 CCACTTCTTT GAATCAAGGC TAAGGAGGGA GCCAGAAGGA AGTATTCAGA
5551 ACACTGAGTA AACTCCAGAA GAAACTACCA TTGCATAAAT CTGGTTGGCC
5601 CTAGGCAGTC TTATCATTCT TGTGTTTTAG TCTTTGCCAG ACTCAAAGTG
5651 CCTATATTTC ATCCCATGAG TCTGCAAACC TGCTTTGTGG TAACCTGCCT
5701 GGCTACTTGC CATTCATTAA CTGCTTCTTG ACCCATGTTG ATTCCCTCTG
5751 TCACTTACTC TGAAAAGACC TGTTAGAAAT AAGCTTGTGA TCTGCTTGAG
5801 ACTTTGGCAA TACTGGTTTA GCCAGAATAG AGAAATCCTT AAGTAGCACA
5851 GCAATCCTTT CTGAATCTTC TATTTGTTTC TTCTTTGTTC TCTGTGTCTC
5901 TCCCACCTAA CATCCCTCTC CAATTTAAGT AATCAAAATA GAAAGAGGGG
5951 CCCAGGCAAG GTGGCCCACG CCTATAATCC CAGCACTTTG GGAGGCCAAA
6001 GTGGGTGGAT TGGTTTAGCC CAGGAGTTGG AGAACAGCCT GGGAAAGATG
6051 GCAAAACCCC ATCTCTACAA AAAATACAAA AATCAGCTGT GTATTGTGGC
6101 ATGTGCCTGT AGTCCCAGCT ACTTGCGGGG TCTGAGACAG GAGGATCACT
6151 TGAGCCTGGG AGGTCGAGGT TACAGTGAGC AGTGACTGGA ATGCTACTGC
6201 ATTCCAGTCT GGGTGACAGA GGGAGACCCT GTCTCAAAAA AAAAAAAAAA
6251 TTTGAGGGAA TATAGGCAGT GCAAGGAAAG GCAGAATATA GGCAGTTCAA
6301 GGAAAATTTC CTTGATACAA GTAGTGTCAA ATGCATATAC ATACATGAAC
6351 ATCAAGAAGA AATATTATTA TTTAAGTAGT CTTAACATGG AGAAGGAATC
6401 TTGTTTTTCA AGAACTGGTC TCTGTGGTCT GCTTAATTTG CAGAAGACAA
6451 AGGCATAATT TGAGATAATA AAGAACAAAG ATAGGTTATT TTCTCAAAGT
6501 ATGTATAATT ACAGTTAATT AGAGACATTT TTGGAATATT GTAGTATTCT
6551 TTGCCTACAA AACTCAAGAT CTATTTCTTT TTATGGGGCA GGGGGGCGTA
6601 GGTGGGTAGT AAACTTAGTT AATGAAGTAA AAGGCGCTAC GACTGAAGAG
6651 CTCTTAAATT ATGTAATTAT GTAAAAAAAG TAAAGCTTTA TTAAATATTA
6701 ATAACATCCG AATGTAGTTA CCAGTGAATC CATTAAGGGC AGATGCTAAA
6751 TTTGCCAGTA ATTAAATAGA GAGCAGAGGA AATGGTGTAT GCTGTGTTAA
6801 ACATAGAAGT TGCCATCTCA AGTAACAATC AGTCTTTCAA AACAGATGGA
6851 CTGAAGAATA TGTTCCAGTC ACCTTCGCAA ATTATTTCTA CTTAATTTAC

FIGURE 3C

6901 ATAATAATGT TTAATGCTCC TTTGTCTAAA TGCTTAATTT TTTAACATAA
6951 GCAGTAAGAG GGAAAATCAC TTTATAAAAG GTTGGGAGGG TGAAGGTGGC
7001 AGTGTTGAAA ATGATTAGGT CTTGCTAGAA AAAATACCTT TATTTTCTTT
7051 GAAAAACACT TATAAGAACT ATAAGAACTA AGGTAATAGT CAGTGTATTG
7101 GTGCTTTGTG TTACAAAGTG TCTTCACATA TTTTATCATC TCAGCAATCC
7151 TTCACAATGA TCTGGGGAGG GCAACTGTAT TAGCTTCATT TTATAGATGA
7201 GGAAACTGAG GTCCAGAATT GCTGCCAAAG CCACAATCTG TTACATGCAG
7251 TGCAGGCTCT TGACTGCATA TATCTCTTTA CTCTAGAAAT TTGCTAACTC
7301 ATTACAACTT GTTTATATTC CTTTCCCCCA ATTCTTGAAA ACCTTGGTTT
7351 AAAGCCTCAA TTGGTGACAT GGGCTTCTTA TTTCCTTGAG GTTTTTTTGT
7401 TTATTCCTTC CTGCAATAGT AGGCTTCTTA TATCCGTTTA TTACCAGGAC
7451 TGAACCTTTC ACTATAAGGG CTATGAAAAT AAGGGGGAAA ATGTTCTATA
7501 AGCTTTAAGT ATGATTTTTT CTAAGCAAAT GTCAAATTCT ATTCTGCATA
7551 ATGTAATTGG ATAAGGAATT GCTTATTTTA ACTCACTTTG AATTGGATTC
7601 ATTAGTATTT GAATTTGGGT AGGATTTATA ACTTTAAAAG CANNNNNNNN
7651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

FIGURE 3D

9201 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
9251 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
9301 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
9351 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
9401 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
9451 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
9501 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
9551 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
9601 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
9651 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
9701 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
9751 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
9801 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
9851 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
9901 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
9951 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10001 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10051 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10101 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10151 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10201 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10251 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10301 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10351 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10401 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10451 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10501 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10551 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10601 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10651 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10701 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10751 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10801 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10851 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10901 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
10951 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
11001 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
11051 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
11101 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
11151 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
11201 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
11251 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
11301 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
11351 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
11401 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
11451 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN

FIGURE 3E

11501 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
11551 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
11601 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
11651 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
11701 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
11751 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
11801 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN
11851 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNA CTTTATCAAA
11901 AAATTGATGG GGAGAGTTTG TTGAAGCTCA GAGTGAGGAT GGATGTAGAA
11951 CATTTCAAGT GCTTCATATC CAGAAAATCA GTAGTCCTCC ATCTGAGTTG
12001 TAGACACAGG AAAGGAGTTG AAGATGAATG GAGTAGGAAT GTAAAAGCCT
12051 TATCTTTACC CTCCTCAGCT TTAGGTCTTA ACAAGAATGA GCCTCCCTTA
12101 GTCTTTCTTT ATGCCCCTGT CCCTGAATGT TGGTGATGAC ATTGTTTTTC
12151 CTGTATTGAA TACAAAAATA TGGCCAGTAA TTTAGGAATC AAGAGGATAT
12201 AATTCGGAAG TAGACTGTTG TGTTTAGGAG TTTTTCTTTC CATTGTGGAA
12251 TTGAGTAGCA GCGGTATATA TGCTATGTCT GGTAAAATGG GCCATACAGT
12301 AGTCTAAGAC ATGAGGAGAC CTTAAGGAGC TTGGACTTAG TTGAGGTGAC
12351 CAGACTATTT AATCTGCTTA GGTGCCACAG CAAAATACCA TAGAGTAGGT
12401 GGTTTAAACA GCAGACATTT ATGATCTCAT AGGTTTGCAG TCTGGAAGTC
12451 AGGGTGCCAG CGTGGTTGGT TCCCGATCAG GGCTCTCCTC CTGGATTGCC
12501 CGTGTCCTCA CATGGCATAG AGAGAGTATG ACAGCATGAG CAAGCTCTCG
12551 TTTTATCTTC TTATAAGAGC ACTGATCCCA TCATGAGGGC CCCATTCTCA
12601 TGACCTCATC TAAACCTGAT TATTTTCCAA AGGCCCCATC TCCAAATGCC
12651 ATCACATTGA GAGTTAAGGC TTCAACATAT GAATTTGGTG GGGAAACCCA
12701 GACATTTCAA TCCATAATTC AGGCAGATAT TTGGGAAGTA ACACAGTTGA
12751 AGCACTGAAT GCTATATTTC GTACTATCTA AAGAATCTAG GATGTAATAA
12801 ATTTAAGATG CTTCATTGCC AATTAAATTA AGATACAATG CTTTTTTGAT
12851 TACTTAGAAT TTTTTAAAGA GCTCTTTTAG AGTTAGACAT AGATTTTTGT
12901 CATATGTCAC TTGCACATTC AATAAGATGG AAAACACAAG TGAAAAAACA
12951 CATAAGGAAT TGCTAAATTT CACATATTTA GAGTCTGCCT TCTGAATTGT
13001 TTTTGGAGTC AGAGTTGTTA ATACCTGTAA TTTTCCGTTA AACATCCTCT
13051 GTGCCGCCAA GAGAATTGGT GATGTAGCAT TCCTTTCAAG ATCCCAAAAA
13101 AGAATGCGAA GGTTTTGGTG CTGGCCTTCA GCTTTGCAAT TATGCAAAGC
13151 CAGCCTACTT TGACTGCTGC TTAGGGATTC CCCATCTTCT ACTTCCTTCC
13201 CAGTCCATTT GGTTCCTAGA GGGTGAAATG AATGCTCCAG TATCATTTCT
13251 GGGAATTTCT TTCAGGCTGT TGACTGTCAT ATGCAAATGT CATGCTGGCA
13301 GTTTTGTTAT TTTCCCATGT GTAAGCAATG ACAACATCAT AATTGGCTTC
13351 TGTCTGATAG CAATTGTAAG AGGAATCCCA ATTTCTGAAA TGTTACCCAA
13401 AAAAGTGACT TTAATTGACG AAGTATGATG ATGTAGAAGG ATAGGCAAGA
13451 AATGCAAAAG GTAATTTAGA AAGGTTTCAT GGGTAAAATG TGACCTATGT
13501 GATCTAGGGC TATAAAGGAT TTCAATAAGC AGAAGCACGA GGTGGGTTGT
13551 TGAAGAAAGC ACTAAATGTT TTTGGATAAA GAATATAATA ATTTGAGAGT
13601 AAAGGGTAGA GGGAGGGTTA TGTAGGTAAG TAGTTGTAAG ATGGGGAAAG
13651 ATTGGGTAGT ATTTAGCATT TATCCTTAAT GTTGACTTCA GTGTAGTTCT
13701 CTTTGTGTGT TTTCTAGTAT AAACTGCATA CATGAAAGTT AAGAATCTTG
13751 TGTTAAGTCC CATATAGGAA GGAAGTAGAT AGGAAAACCA AACTGGAAAA

FIGURE 3F

13801 ATGTATGGAG ATGTTGGTGA AATGACAGGA ACGAAAGCAG CTTGTCTGAG
13851 CTTGATCTCT TCACTTCCTC AGTGGTGGTT CTGAGCGCTG GTTTGGCTGA
13901 ACTCCACTTA CCAGGGAAAA GGGCATAAAG TAAACAGGGT TTGTGTGGAA
13951 GAAGTGGAGT AGAACAAAGT GGAGAGGATC TCTGTTCATT TAGTGTATCT
14001 GACAGTGTGC TTGTCAAGTC ATAAAACACT TGAGGATGGA AATCTGGAAG
14051 TCATTGTATA CATTTTCTTC TTTCCCTAAC ATCTAGTCAG TTACAGTTTC
14101 TGCCAGTTCT TTTGCTTTTT CCATGTTTTT GGAGGCTGTT CCTCTTCGCT
14151 CCACATGTAG TAAATGCTCT AGTTCATGAC CCATGTCTTA TCTGGACTGC
14201 CATGTCAGCT TCCTAACTCA TCCATTCACA GCACCAGTGA CTGTAAAACA
14251 GCATTAGTGA GGATAAAACA GTGGCTGTCA AACTTTTTTG ACTGTGGCCC
14301 CCAGTAAAAA TACACTTTGT ATTGCAACTT ATGTATACTT TATATATGTA
14351 TGAATAATTA AAACAAAAGG TTGATTCAAG AAAAATCTTT ACATTTACCC
14401 TGTGCCATGC AATCTTATAT CTTGTATTCT TTTCTGTTTC ATTTTTTTAA
14451 ATGTGTGCTT GCCATCCACT AAATTGATTC CGGAGTTGGA AAAACACTGA
14501 CCTGACAACT AATATCACCA TGTTATTCCT TAAACTCTCC GATGGCTTCT
14551 TACTATCTTC ATGATAAATT TGAAGCCCTC AACATCAGCA TACCAGAACC
14601 TTCATGACCT AACCCTTACC TAGTTATTCT AATCTATTAT TTACCTGATC
14651 CACTCAGCTC ACATTTCATT CCAATAGACA AGTAAAGTTT TTTGTAATTC
14701 CTTGTAGCTT GCCTTTCTTC ATGGTGTCCA CTCTGTTGAA AATCTACTAC
14751 CCTCCATTTC TTCAGTGCTT TACTGCTTAC TCCTACCCAT TCCTGGGGCT
14801 CAAGTCAGGC CCCTATAACC AGGATGCTTT TCCTAACACT CCTTGCCCTA
14851 CCACCAGGCT GGGTTAGGTA GTTCTCCATT ATATAATGTG GTTCTCAATG
14901 TTGTTACCTG TTTATTATTA TGTGTTTTTC TCTTATTGTC CCATAAAATA
14951 GTGAATATTC GAGAGGATAA GGAAGTCTCC CATTAAGCAT CCCTAATGTT
15001 TAGTATGTAA CATGTTGGCA TTGGTTGGAT GAATGAGAAA AAAAAAAGAT
15051 TCTTCTGTTT GGAAGGAAGA TACAACTGGT ATCCCTTAAG TCTTTTCTTT
15101 TTTTTTTTTT TTTTCCTTTC TCTATAGACA AGGTCTCACC ATCACCCAGG
15151 CTGGAGTGCA GTGGTGCAAT CACAGCTCAC TACACCCTTG TACTCCTGGG
15201 CTCAAGTGAT CCTGCTACCT CAGCCTCCCT AGTAGCTGGG ACTGCAGGCA
15251 TGCACCACCA TGCTCAGCTC ATTTTAAAAA AATTTTTTTT GTTGAGACAG
15301 AGTCTTGCTA TGTTGCCTAG GCTGGTCTTG AACTCCTGGG CTCAAGTGAT
15351 CCTCCTGCCT CAGCCTCCCA GAGTGCTAGG ATTATAGGCA TGATCCACTG
15401 CACCTGGCCC CTTAAGACCT TTAATTGCAG AGCAGCAGAG GACAAATGAC
15451 ATAAATACAG GATTTGACTT TCATTTTTAA GTATCAAATT AGTGATGGGT
15501 TGACAAACAA GTCATACAGA ATGTTCATGA ATCAGTTCGG CCCAGGTAAC
15551 TCATAACCCA AGACCTTTGG GTCAATGAAA TTCTGCCACC TAAGTAGCAC
15601 CATCCAATGA TGTCATACCT AAAAAGGAAA TTGAGTTGTA GAATTTTAGG
15651 TTTTAGGATT CTTTCTCTAA AACTGAGGAG CTGTGCCACT CTTCAAAGCC
15701 TCACAATTAC ATTTCATTGG TTCTTATGCC ATCTGGGTTC TGGTTAGAGG
15751 GCTGATGGAA GTACTCAAGA AATATTGGAA GTACTCAAGA AATATTAGAA
15801 GGTGGGAAGA AGGTACCTCT CTTGTTCTTG TCAGTGGCAG CACCAACAGT
15851 GGGACTTTGG GTCTCTGGGT TCCAGCTCAG CAGCAGAGGT ACTAGTACTG
15901 TAGCTCCAGC AGCTTCAGCA GGAGTGCAGG CTCATGGGAT CAGAGAACCA
15951 CCTTTTCCGC TTTGTTCTTC CAGCCCAGCC AACAAGTTTG TAGCTATTTC
16001 CCTGCATTAA AACTCCCCTC TGTTTGAAAT ATCTATAGTA ATTTTTCTTT
16051 TCCTGACTAA TACAACCTGT TAAAGAAGCT GAAGCTCTGG TAAGTTAAAT

FIGURE 3G

16101 GCCCAACAAT GGTCTTGAGT AGCTAGTGAT TTTTGTTGCT ATTGGTAAGT
16151 AAATCTAGAC ACTACTTTTT AGTCCCTTTT TTAAAAGAGG ACTGGTTTAT
16201 CTATGATGAA TACATGATTG ATTGATTGAT TGATTGATTG ATTTTTACTT
16251 TTTCTTTTTT TTTTTTTGAG ACGGAGTCTT GCTCTGTCAC CCAGGCTGGA
16301 GTGCAGTAAC ATGATCTCTG CTCACTGCAA GCTCCTCCTC CCGGGTTCAC
16351 GCCATTCTCC TGCCTCAGCC TCCTGAGTAG CTGGGGCTAC AGGCATCTGC
16401 CACCACGCCC GGCTAATTTT TTTGTATTTT TTGTAGAGAC GGGGTTTCAC
16451 CATGTTAGCC AGGATGGTCT CGATCTCCTG ACCTTGTGAT CCGCCTGCCT
16501 CAGCCTCCCA AAGTGCTGAG ATTACAGGCA TGAGCCACCA CGCCCGGCCT
16551 AATTTATTAA AACTTTCGGG TGGTCAGGTA ATTCTGATTT GTCAGCCATA
16601 TTTCTAAATT ATCAATNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
16651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
16701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
16751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
16801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
16851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
16901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
16951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18101 NNNNNNNNNN NNNACAGGCA CACACCACCA TGCCTGGCTA ATTTTTTGTA
18151 TTTTTAGTAA CAGGGTTTCA CCATGTTAGC CAGGCTGGCA TCGAATTCCT
18201 GACCTCAGGT GATCCGCCCC CCTCAACCTC CCAAAGTGCT GGGATTACAG
18251 GCGTAAGCCA CCATGCCTGG CCTGTATTTA ATCTTCATAG CAGTTTTATG
18301 AGGTAGGTGG TGTCATCCCC ACTTTACAGA GAAGTGGGTT AATGTAGGGT
18351 TCAAATGATA AATAGTAACT TGCTGATAGT CACTGGCAAT TTTAATTTGT

FIGURE 3H

18401 CTTCAGTGTA GTAGAGTAAC TGTGAACTGT TAGAGTTATG AAACTGACAT
18451 GGAAAGTTGT ATACCAAAGG AGTCTTAGGA CTGTCCATGG ATACTGTTAT
18501 GTATCATTTC ACTTATATTG GCTTCAGCTT GCGATTTCTC TACTGTAAGT
18551 GGTGAGAATT GATCAGATAG TTAAGGAAGG TCCTTAGATA ATGCAGTATA
18601 CTTATTAACA TACAGACATC AAGAAGCAGA AATATATAGA CATCTTCCTT
18651 TTTGGTTCTA ATAGGGCTTC GTGGGACACA TATGCAACAT GCCTATGATT
18701 TTTACAAGCC TGATATGCTA TCTGAATATC CTATAGTAGA TGGAAAACTC
18751 TCCATACAGT GCTACCTCAG TGCATTAGAC CGCTGCTATT CTGTCTACTG
18801 CAAAAAGATC CATGCCCAGT GGCAGAAAGG TAAGTTTTAC CCATTTTCCT
18851 TGGTTTTGGT ATGAGTTGAG AGCAGTCTAA TGTACTAGGT ATCTTTGGTA
18901 GGCAACTACT TTGTGGGCAT TCTTCATTTA ATATCCTTTT ACCATTAATT
18951 CCTCATTCAC CAAACAACAT TTTCCCATAG TTTCTGGGAA AGTGTAATTT
19001 ACTAGAAGAG GTAAACTTTG GAACTGAGGT GTATCTCTGC AAAAATATTT
19051 AGGTCGGTTT ACCCCTTGTA AGAAAATCAA AGTGGAGAAA AGAAGGTAAG
19101 TTGAATTTTG TTCATCTTTT GAGAGAGGTA TTTTAACAAG GTTTTGGACT
19151 ACAGCTGTGA TTCAGGGAAA GCTAATGAAA ATGAATTACT AAAGTGATCT
19201 TACCCCAAAA ATAATCTTTT TGCACTTGAC CTGTGAATTT GTATTTGTTT
19251 TTTTACTGTT ATCATTAATC TGGAAATTTG TTGAGGCACT GAAAGGACAG
19301 TATTTGAGTT AATGCTATCA TAACACATTA TTACATAAAG TATACTTTTT
19351 CTGTAGTCCA ACTTTGCTTT TTAGAGGTTA TGAGAAGGGG TTAAAAATCA
19401 TATTCAATGA CAAATATCAG TGAATTTAGT CGCTCTGGAT AAGAAGCATT
19451 CTTGCAGTAT ATATTAACAG AATAGTGGTT TTCTAACTTT TTTATTAGGA
19501 CCCACAGTAA GAAGTACATG TTACATTGTA TGTGTATGCC AGACTGAAAC
19551 AAAAATGTCA TGACATTACT TACCCTTGCT GCAAGTTATT CAGTTTGCTA
19601 TTTTTCTACT GCATTTTGTT TTTTAAAATA CTCTTTTATT TAAAAAAAAT
19651 ACTAATCCTG ACCCACTAAA TTGATTATGT AACCTGCTAA TGTGTATGAA
19701 TCTTAAATTT GAAAATTAGT GACATAGTAC ATATTGTTTC ATCTTTGAGT
19751 GTCTTTTTAA ATGTATACTT TAAGGTATAG AGAGGTTTCA TTATACAGTG
19801 TATTTGTGGT TGCTGTTTAA ACATATACAA ATATCCTAGC TTTATTCTAA
19851 AGTCAAACTT TAAAATTTCA TGGCTTATAT GAATTTCATA GTTTCCTTGG
19901 ACTTCTCTTT CAGAGGGAAA TGATAAAGAT TTTACCTTGA ATGATTTTGG
19951 CTTCATGATC TTTCACTCAC CATATTGTAA ACTGGTTCAG AAATCTCTAG
20001 CTCGGATGTT GCTGAATGAC TTCCTTAATG ACCAGAATAG AGATAAAAAT
20051 AGTATCTATA GTGGCCTGGA AGCCTTTGGG TAAGAGGAGC TATTATGAGT
20101 TTTTTCCTTC TATATTAGAG CATTTTTAAT ATCTGTTAAG CTGTTATTTG
20151 TACAGACCTG AGAAATTGAG AGTCAGAAGA ATCTTAGAAG TCATCCAGTC
20201 TAATCTGTGT GTCTCAGTCA GTGAAGAATC TAAGTCCAGA GAGGTGGTAG
20251 TTAACATGCA CAAATTCTTT AGACATTTCT ATTCAGATTT TCTGATTTAT
20301 TTCTTTCAGC TCCATTCATG TTGTCACGAT AAAGTAACTG CACAAGGGCC
20351 TATATTCACT ACAGCAGCCT CTTAACTCCT TACCTCTCTC AGCACCCCTG
20401 CCCCCATGCC CTTTTCCATC CTGCACACTG CCACAGCTAA AGTCAGCTTT
20451 TGTACTCCAC CTGTCTTTTT CTCACTTTAG GCTCCCTAGC ATGCTATGTG
20501 TGTTCAACTC GTTCTGTTTC TCCCTGTGTC TCTTGTGTGT CCTTTCTCTA
20551 TCTGATAAAA TTATACTTGA CTTTTAAAAC TTGGCTCCTG TAATACCATG
20601 ACTTTTCTAA CTAAATAAAC ATTATTATGG ACTTGAAATA GTATTCTATT
20651 CAGTTGATGA ATATTCAGTT GATTGAATAT TCTATTCATT GAAGCCAATA

FIGURE 3I

20701 TAAGTGAATA TAAATATAAA GCTACAGTGC GTCTTTTAAC CTATTCAAAT
20751 CAAGCAGGCT TAACTTGATT ATGAAAACTT TTGAGAAAAA GAACCATATA
20801 TATACAACTG TTATGATTTC TATAGCAATT AGATTGCTGC TACTTGGCTT
20851 TTAATAAATG AGAAAACAAT TATATACACT TAAAGATTTG AATCCTAATT
20901 AGGCCTGCTG TTTAGTGTAA TAAAAACATA GGCTTTAAAC ACTGTAAAAC
20951 TGTAAAATAA ATCTTTCAGG GATGTTAAAT TAGAAGACAC CTACTTTGAT
21001 AGAGATGTGG AGAAGGCATT TATGAAGGCT AGCTCTGAAC TCTTCAGTCA
21051 GAAAACAAAG GCATCTTTAC TTGTATCAAA TCAAAATGGA AATATGTACA
21101 CATCTTCAGT ATATGGTTCC CTTGCATCTG TTCTAGCACA GTAAGTATAA
21151 ATTTCACCTA CTACTTAACT CCCCTTATTT GGGAGATGTT AGATTTCTAA
21201 GACCAAATCT AGTGTCAAGC ATGTTGGTGG TAGATCACAG AAAATTTTAT
21251 CTTGAGGCTC TCTAATCTGC TATTGTCCAT TGACTTGAAA GATGTATGGG
21301 TTGAGGCTAC AGTTCTTCCA GAAGTATTTG TTAATTTCAT ACTGGCTTTC
21351 CTGGCTTCTG TTTTCATGGT TTTTTAATTC TTGACCTACA GTTGAACCAT
21401 AAATACCTGG TTGATGAAGT AACTTGTTTT GTGGCATGAC TTTCACAAGC
21451 TCTGTCATTC CCCACAAGAT GAAAACTCAC ATGCTGCAAT ATTAAAACTA
21501 AGTTATATTC CCTACTGCAA TATTAACACT TTGAGTTAGA TCCTTAAAAC
21551 TTTAAGTTAG ATTCTACTTT TACTTATAGC CTAAATTTTT ATTGCTACTT
21601 TTATAGCTTC CCACACGCTG TAGCTTTGGA TCAGTTAAAC TTCTGAACTA
21651 TTGTTACACC CTACATAGGT ACTCACCTCA GCAATTAGCA GGGAAGAGAA
21701 TTGGAGTGTT TTCTTATGGT TCTGGTTTGG CTGCCACTCT GTACTCTCTT
21751 AAAGTCACAC AAGATGCTAC ACCGGGTAAG TGCTGAATCT TTCAACAAGA
21801 ATGTATTGAG AACTGAGTCC AGGCACAGTG GCTCACACCC GTAATCCCAG
21851 CAGTTTGGGA GGCCGAGGCG GGCAGATCAC CTGAGGTCAG GAGTTCGAGA
21901 CCAGTCTGGC TAACATGGCT GAAACCCCAT CTCTACTAAA AATACAAAAA
21951 TTAGCCAGGT GAGGTGGTGC ATGCCTGTAG TCCTAGCTAC TTGGGAGGCT
22001 GAAGTAGGAG AATCACTTGA ATCCAGGAGA GGGAGGTTGT GGTGAGCCAA
22051 GATCACACCA CTGTGCTCCA GCCTGGGTGA CAGAGCGAGA CTCTGTCAAA
22101 AAAAAAAAAA AAAAATGTAT TGAGAACTAC TCTGGGGAAG TTGATTTAGC
22151 AGTCTTCTCA AGTGAGCACC TGAATCTGTC CCACAGATCA TTACAATATT
22201 TTAGTCTTCA TTACTTCTTT CAGTAGGTTT TTACTCTCTG CCCTAAAAAT
22251 CTATCCAAAA AAAAAAAAAA ATTCTACCTT ATCTGGATAA AGGATAGGAC
22301 TAAGTTATCT AATTTTTATA GGCTTATGGT CTTGGCTATA TTTAAGGTCA
22351 CTTTTGTGCT TTCCCTGAGC AGGAAAGAGC AAAAATGTAG AGATAAACTG
22401 ATGAAAACTT GACATTACTT TTTAAAATTA TACCATGGGC CAGGTGCAAT
22451 GGCTCACACC TATAATCCCA ACACTTCAGG AGGCTGAGGT GGGAGGATTG
22501 CTTGAGGCCA GATGTTCAAG GCCAACCTGA GCAACATAGT GAGACCCCAT
22551 CTCTATAAAA AATAATAAAA ATAAAATAAT TATACCATGG ATTAATTGTA
22601 GACAAGTTAT TTATAGTTTC AAATTATGCC TGTTTCCTAA CTTGTCTAGT
22651 GGCAGATACT CAATAATAGA TTTCTAGTCT GACATCATAG GAGATTTGTC
22701 AAATAGGTAT CATCTTATCT TTTAACTAAT CAGTAGCCAG TAGTTTTAAT
22751 GAAAATGAAA AGTTGTTTTG CCTCATTTGG CAACATTTTA CTTAGGCTTC
22801 TTTTGGACAT GATTTTTCAA AAAAATCTTT TAATGTTGAA TTATTCACTA
22851 TTTTAGGGTC TGCTCTTGAT AAAATAACAG CAAGTTTATG TGATCTTAAA
22901 TCAAGGCTTG ATTCAAGAAC TGGTGTGGCA CCAGATGTCT TCGCTGAAAA
22951 CATGAAGCTC AGAGAGGACA CCCATCATTT GGGTAAAAAT ATTAAATGTT

FIGURE 3J

23001 CTTTAAGTTA ACCCATTTGG AGGGCTGATA TCATTAAGGA TGCTACATAT
23051 ACGATAAGGA TATCAAGACT TTACTCAGTA CTAATCTGAT GTCAGTGAAA
23101 ATTATTGGGA TATATGAAAC TTATCTTTAG CTTTATTACC AGATGAATTG
23151 TATATCATAA CTAATTGTAG ATATTCTCTC CCTTTCCTTT AGTCAACTAT
23201 ATTCCCCAGG GTTCAATAGA TTCACTCTTT GAAGGAACGT GGTACTTAGT
23251 TAGGGTGGAT GAAAAGCACA GAAGAACTTA CGCTCGGCGT CCCACTCCAA
23301 ATGATGACAC TTTGGATGAA GGAGTAGGAC TTGTGCATTC AAACATAGCA
23351 ACTGAGGTAA ATAAAAGAGT TCCCATCTCC ATATCTTAGG GTTTAGGAGA
23401 CCTAACTGGG ATTTAGCAAC ATAAATAAAT GTCAGTAAAG AAGAGTAAGG
23451 GCTCTGGGAG TAGATTCTAG CTGTACTATT TCCAATTGTA TAAAGTGCTT
23501 TGCATTTGAA TTATTAATAT TTTAAGAATA TACAGTAAAG GCCGGGTGCG
23551 GTGGCTCACG CCTGTAATCC CAGCACTTTG GGAGACTGAG GCAGGCAGAT
23601 CACGAGGTCA GGAGATCAAG ACCATCCTGT CCAACATGGT GAAACCCTGT
23651 CTCTACTAAA AATACAAAAA TTAGTTGGGC TTGGTGGCAC GTGCCTGTAA
23701 TTCCAGCTAC TCAGGAGGCT GAGTCAGGAG AATGGCTTGA ACCAGGGAGT
23751 CAGAGGTTGC AGTAAGCTGA GATCACACCA CTGCACTCCA GCCTGGGCGAC
23801 AGAGCAAGAT TCCATCTCAA AAAAAAAAAA AAAAAAAAAA AAGAATATAC
23851 AGTAAATACT AGGTTTTATT AATGATACCA GGATTTAAAG GAAGACTGAT
23901 ATAGAGAGAA GGTTCATTTG TGGTGTGTGT CTTTGTGAGA GATGGAGTAG
23951 AGGGACAAGG ATCCTTTCAC ATCTCATCCC AGATCATGGT CAAAATCTGT
24001 CCTCAAATTG TCAAGAAGTA ACAATCATAG CTATGATTTG AATTCCTGTT
24051 ACCTGCTAGG CACTTTACTT ACGTTTTCTT ATTTAATCCT TACAACAACC
24101 TCCTTGAAGT TTATAAATGA TACTGTCCTC CCTTTAGAGA TGAGCCTCCA
24151 AGAAGTTACA TTACTTGCCC AGGATTATAG GTAGTAAGTA TTAAAGCCAG
24201 GTTATAAACT AAGGACTTTA TAACCTTGAA ACTACTTATT TATCTGCTTA
24251 CTACAAGTTT GGTAAATGGA TAGTCTTGCT TTTTGCTATT ATACAAATTA
24301 GGTAGCAAGT CAAACCGCCA CTGTTTGAGT TGCAAATACA AGACGTAACA
24351 AGTAAAATAC TGTTACGTGG TGGGTCTCTG TGGCAGGCTT CCTCTCCCCC
24401 CCATATGGAT AATTGTATAC TAAATTCACC ATAAGGTGAA AAATGGATAT
24451 TGAGTTCCCT TCATGAAAAG TTATATAAAA TATATATTTA GCATAAACTT
24501 CTCCAGAGTT GTCCTTTATT AAGTTTCTTT ACAGAAACTT TAATTGGTGC
24551 CATGATTCTT GTGGGGGAAA GAATCATAAG AGCCATCAAC TTTTTTCCTT
24601 TCATTTTAGC ATATTCCAAG CCCTGCCAAG AAAGTACCAA GACTCCCTGC
24651 CACAGCAGCA GAACCTGAAG CAGCTGTCAT TAGTAATGGG GAACATTAAG
24701 ATACTCTGTG AGGTGCAAGA CTTCAGGGTG GGGTGGGCAT GGGGTGGGGG
24751 TATGGGAACA GTTGGAGGAA TGGGATATCT GGGGATAATT TTAAAGGATT
24801 ACATGTTATG TAAATTTTTA TGTGACTGAC ATGGAGCCTG GATGACTATC
24851 GTGTACTTGG GAAAGTCTCT TTGCTCTATT TGCTGACATG CTTCCTGTTG
24901 TGGTCTGGCC AATGCCAAAT GTACTCGAAT GATGTTAAGG GCTCTGTAAA
24951 ACTTCATACC TCTTTGGCCA TTTGTATGCA TGATGTTTGG TTTTTAAACA
25001 TGGTATAATG AATTGTGTAC TTCTGTCAGA AGAAAGCAGA GGTACTAATC
25051 TCCAATTAAA AAATTTTTTA ACATGTAAGA ATTTTGTACT TTGAACAACA
25101 AGATTACAGA AAGTACCTGT GGTTTTTGGA AAACATTTCT AGCTTGGGGA
25151 ATGTGACAAC ATTCCCCAGT GTGGTAAAAT TGGGGTAAAA TGTGGTAAAA
25201 TGTGATACGC ACAAACCCTT TGAAAATAGC AAAACAAACA TGCCCTTTTT
25251 CTAAAATTGA TAAATCCTAA AGAGGAAGAA AAGAGCTGGG ACAATAAAAC

FIGURE 3K

25301 ACTGGCTCTG GAATCTGGAA TGTTAAGTCC AGGCCAGCAG TGACAAAAGT
25351 TATTGTAATG ACCTCTGAAC AGAGAAACAC TGCCATTGAA GAGGCTTCTG
25401 GTATAGAAAA CATGGTACAT TCAGGAGCTG TGAATATAGC TCTAGGTGTG
25451 CTCCTGAATC AGTTCATGGT AGATTATGCT GAACAACAGT GAGATGTTAT
25501 TGGAGGTGTG GATGAGGGAG TTTGTTGTTG CAGTCCTTCT TTGCACCTTA
25551 TTTTAAAGAA TAAATGAAAC ATTTTTCTGG TTACTTTTTT AAAAATTTAA
25601 AATGGAAGGG AAGAATAGGG GCAGGGCATT ATTAGGCTAT TTCTGATGCT
25651 TCAGTGTTAT AAATTCAACA TAGAGGCTGA CAACCTAAAT TCATGGTGTA
25701 ACACAGCTCT TTTCCTTTTC CTTTTTTTTT TTTTTTTGGT ATCTGTTCAA
25751 TGAAAATAAG GTATGACCCA AGTTTTTACC TAGTCTGACT AGAAGTATTC
25801 CACTTCAAGG TCTGAAGTAG GACTTTTACC TTAAAAAACA ACAACAAACA
25851 AAACTATCAC ACAGGATAGA TAAGAAGATT GGTTAAACAG TTTTGTGTAG
25901 ATCTTTTTGG TGCTGAACTA TGACATGAGC CTTATAGATT GTAAAATAGG
25951 GATAGTTGGA ACTAATGTAC AGAACTAAAT TTTTTAAACT TTATTTGCTG
26001 TTAAATTCTG TGAAGTTTCA GTTATCTAAA ATAAATATAC ACAAATATGA
26051 AATATAATGT TTCAGATTGC AAGGTAATAT GTAATAGTAG TGTTTGTAAG
26101 ATACTCTTGT CTAATATTAA CTAGTAGTAT TTTGATTTGT ACAGTCATAA
26151 TTTGTTAAAA TGACTTCATT TAACATTCAC TGATGTAGAT TAATAATGTA
26201 AGTTCTGATT TAAAGAATGG TGGCAAAATG GTGCATGTAA TACTTTTGCA
26251 AGTGTTGGGG AGATCGGTAT GTTTTGAAAA GAGTAATTTA ACTTTTGGGT
26301 GCCAGGAAAT GGGTTTTCTC AAAGTCCATT GCCGGCAATG GGCAGGCCTG
26351 CAAATACTGG CACAGAGCAT TAATCATACA CCTTATTAAC GGTGAGGTGA
26401 ATAACTTTGA AATAAAGTTT TAGAGAAATG TTTCAGATAC TTGAGTATTC
26451 TTTTTCACTC TTGAACTAAC AACTTCGGCA AGAAATCAGC TAATATTCTA
26501 TTTTTAAATA TGGGCATTAA TTTCATTTCA GTTCGTTCAC TCATTCCATT
26551 CATTTATCAT TTCACAAACA TTTGAAATCC TAATATAAGC AAGGTGCTCT
26601 GTTTAAGGCA GAAATTTGAA AATGTACAAG ATATATGGTC TTGTCTTTAA
26651 GGAGCTGTTC ATCTAGAATG GAGGAATTTA CACTGATAAT TATTCCTACA
26701 CTTGAAACAA AGAAATTAAC TCTCAAATTG CGTGGCAAGC ATATATAGAC
26751 TTTGCTATAA ATATTTATGA AATGAGTTAC TGTTTTCCTT AAAAAAGCTA
26801 AGACTAAGGG CTGGCAATCA AATAAGAGCA AATTTAGTGG TGAACGTAGA
26851 ACTGCCCACT ACCAGCTAGA GTCTCCAACC TAAAAGTCCC ATGTTGCTAG
26901 TGATCCCCAG GGGTTTTATA GAAGGAATCC CTGCATTGGC AGTAATTTTG
26951 GATTAGATGA TCCCTAAGAG CACCATCAAG TCTTAGGATT CTATGAATTA
27001 GGAAATAAAC CAAATTATAT ATTTTCTAAT ACTGATCAGC TCATATTTTA
27051 TCATCATGTC ATGTCTGGCT TTCATACTGG GAATACAGAT ATAGAAGGAA
27101 TCGACACAAC TAATGCATGC TATGGAGGCA CAGCTGCTGT CTTCAATGCT
27151 GTTAACTGGA TTGAGTCCAG CTCTTGGGAT GGTATGTTAC ATGCCTATTC
27201 CCCGCCGTCC CCCAAAATTT TTTTCTAAGG TTCAATAGAC CCAAATGACA
27251 CTTTAATTAA TGCAATACGC AAACTTTTGT AATTTATCCT TGTTTGGATA
27301 TATTAAGAAA GATATTTTAC CTGTCTGTCA TTATCCGAAT TGTGAATTGG
27351 TTATCTTATC TTGTAGGACA AATGGTCTAT TCAAAATTTA GTCAGATGGA
27401 TGACAGAGCC TTGGCAGATG AATTTTAAAA AAAAATTAGA GCATTTTCTT
27451 TCTTTATCAA AGAAGGGAAA AGCATATTCT GGGGAAAATA TAACAGACTT
27501 CAGTTTCCAT GTTTGGTTAT AGTGTTGAAT TCCTTCTTGT GAAATAACAA
27551 AAAATATTTT TCAGGACGGT ATGCCCTGGT AGTTGCAGGA GATATTGCTG

FIGURE 3L

27601 TATATGCCAC AGGAAATGCT AGACCTACAG GTGGAGTTGG AGCAGTAGCT
27651 CTGCTAATTG GGCCAAATGC TCCTTTAATT TTTGAACGAG GTAAGTGCTT
27701 GGGAAAGCAT TTTTGTTTTT TTTAGCACAA TATGCTGAGA AATTTGAAAA
27751 TAGAAGTAGG AGCTGTCGCT TACTTAATGG TCATTAAATG CAGGTACTAC
27801 TTGCTAAGAG CTTTATGTGT GTTATCATAT TTATGTTTTT TTTTCTTTTT
27851 TTTTTTTTTT GAGACCGAGT TTCACTCTTG TTGCCCAAGC TGGAGTGCAA
27901 TGGCACGATC TCGGCTCACT GCAACCTCTG CCCCCAGGTT CAAGTGATTC
27951 TCCTGCCTCA GCCTCCTGAG TAGCTGGGAT TACAGGCACA CACCACCATG
28001 C (SEQ ID NO:3)

FEATURES:

Exon:	16553-16577
Intron:	16578-18664
Exon:	18665-18829
Intron:	18830-19913
Exon:	19914-20079
Intron:	20080-20969
Exon:	20970-21140
Intron:	21141-21668
Exon:	21669-21775
Intron:	21776-22856
Exon:	22857-22982
Intron:	22983-23192
Exon:	23193-23356
Intron:	23357-24609
Exon:	24610-24696

CHROMOSOME MAP POSITION:

Chromosome 5

ALLELIC VARIANTS (SNPs):

DNA				Protein		
Position	Major	Minor	Domain	Position	Major	Minor
2061	G	A	Beyond ORF(5')			
3388	C	T	Beyond ORF(5')			
4147	-	T	Beyond ORF(5')			
12272	G	A	Beyond ORF(5')			
12936	A	C	Beyond ORF(5')			
13560	C	A	Beyond ORF(5')			
14127	T	G	Beyond ORF(5')			
18789	T	C	Exon	50	G	G
18901	A	G	Intron			
22095	G	A	Intron			

FIGURE 3M

22257	-	A	Intron
22582	A	G	Intron
25232	A	T	Beyond ORF(3')
26224	C	G	Beyond ORF(3')
26695	C	T	Beyond ORF(3')
27982	A	G	Beyond ORF(3')

Context:

DNA

Position

2061
CGTATCTCGCAGCTCCGTCATTGGCAACTGGGCTCTCGTGCCACCTCACGTCAGTCTCTC
ACACCACTTCCTCGGCCCTGAGACTTTGTCCCCGCCTCTTCTCCCCGCCCTTCCAGCCAC
GAGGGAAAATCCTAGCGAGTCATCGCCTCTAGTTTCCTTTTGATTGGTAGAAGCCGGACT
GGGGGGCGGGCGCTGCCGGGCAACTCTACCGGCCGCGATTGGCTGTGGGAGCCACCGTCC
CGCCTCCCAGCGACTCTCGGCGGTGCCGGAGTCGGGTGGGTTGGCGGCTATAAAGCTGGT
[G,A]
GCGAAGGGGAGGCGCCGCGGACTGTCCTTTCGTGGCTCACTCCCTTTCCTCTGCTGCCGC
TCGGTCACGCTTGGTGAGTGTCCCGCGCTGGGGAGTAGAACTGGGCTGCGGAGGTGCCGC
GGGCGGGGTGTGGGCCAGACAGAGGCGGTGTCCTTGACTAGGCCCGAAGGAGCTGGGGCT
CTGGGTCAGGACGTAGGCGTGGACTTTGCCCGGGAGGATGGGGCACCGTGAGCGGGCCG
GGCGGGGGTTCCCTCGTGAGGGACCTGAGGCCGACCGTAGCGGATCTGAGAAGATCCGAG

3388
CCTATGGCCTTAGCCTTAAGGCATCAGTAGCTGCTTTCACTGCTCACCTCTGCTGCAGCT
CCCCACCTTCCCGAGGATGCTCTTGCCACCTGCTGCAGTAGGATGATGTGTTCTGGTTGC
TGCTAACTAACATTTGCTCTGTTTTAGGCATGAATATGAAAAACAATGACAAGATAAACA
ACAAAATTAAGACAAATGGAAGTGCTCCTAGAGTTAACAGATTTTTCCTTCTGAGATGTG
TTTTGGACTTTATTGCACAGATACTATTAGATGAGAGGCAGTTGAAAGTCGTTAACATTA
[C,T]
CCGTGTCAGTAGTTCTTTGCACTTGAGACACCTAAGCAGCTTGTGTTCTTTAAACTTTAT
TTTAAAATTGCAGTTATTTTTGTGTGAAGAAGGGGGCAGGGATAGCATACCTTATGGGAA
GAGAGAAAGGCTTTCTTTGTGTCTACCTTTGTAGATATTTCTCACCTAAGTTTGTAAGTT
TGCCCTTTATTCGGTTCTACTTTAGTTCAGCTCAATTCTAGTATAATCATCAGTAACCCC
AGCACTCAGAAGGTCTGACTTACGCTGTGGGGAGGGAGTGTAAAAGGATATTTTATGTTT

4147
AAAACCCGATATTGTTGTAACATCAGTTTCCTGTGTCCTCTAGAATCATTTAAAACCTGG
TTGGATCTTCCATAATCCAGTGGAATTGGATATGAGATGTAGCTGGAGAAGTTTGTTTTG
CTACATATCAGAATCTCCAATTAGTTTCATTTAGAAAGGAATATAGCCTTATAATTTTAT
GCTGGGTTACTGTGGAACCAAATATCATAGAAGGATGTGTGATATTTTTATGTTTTTCAA
GAAGGTAGTATAGATTTAAAAGGTGGGATACATATTACCTGTCCTAATGATAGGACTAGA
[-,T]
TTTTTTTTTTTTTTTTTTGGGGAGACAGAATCTCGCTCTGTCGCCCAAGCTGGAGTGCA
GCAGCGTGATCTCGGCTCACTGCAACTTATGCCTCCCAGTGATTCTCCTGCCTCAGCCTC
CCAAGTAGCTGGGACTACCGGCATGTGCCACCACACCCAGCTAATTTTTTTGTATTTTTA
GAAGAGATGGGGTGTCACCATGTTGGTCAGACTGGTCTTGAACTCCTGACCTCAAATGAT
CCGTCCGCCTTGGCCTCCCAAAGTGCTGAGATTACAGGCGTGAGCCACCATGCCTGGCTA

FIGURE 3N

12272 AGAAAATCAGTAGTCCTCCATCTGAGTTGTAGACACAGGAAAGGAGTTGAAGATGAATGG
AGTAGGAATGTAAAAGCCTTATCTTTACCCTCCTCAGCTTTAGGTCTTAACAAGAATGAG
CCTCCCTTAGTCTTTCTTTATGCCCCTGTCCCTGAATGTTGGTGATGACATTGTTTTTCC
TGTATTGAATACAAAAATATGGCCAGTAATTTAGGAATCAAGAGGATATAATTCGGAAGT
AGACTGTTGTGTTTAGGAGTTTTTCTTTCCATTGTGGAATTGAGTAGCAGCGGTATATAT
[G,A]
CTATGTCTGGTAAAATGGGCCATACAGTAGTCTAAGACATGAGGAGACCTTAAGGAGCTT
GGACTTAGTTGAGGTGACCAGACTATTTAATCTGCTTAGGTGCCACAGCAAAATACCATA
GAGTAGGTGGTTTAAACAGCAGACATTTATGATCTCATAGGTTTGCAGTCTGGAAGTCAG
GGTGCCAGCGTGGTTGGTTCCCGATCAGGGCTCTCCTCCTGGATTGCCCGTGTCCTCACA
TGGCATAGAGAGAGTATGACAGCATGAGCAAGCTCTCGTTTTATCTTCTTATAAGAGCAC

12936 CCATCTCCAAATGCCATCACATTGAGAGTTAAGGCTTCAACATATGAATTTGGTGGGGAA
ACCCAGACATTTCAATCCATAATTCAGGCAGATATTTGGGAAGTAACACAGTTGAAGCAC
TGAATGCTATATTTCGTACTATCTAAAGAATCTAGGATGTAATAAATTTAAGATGCTTCA
TTGCCAATTAAATTAAGATACAATGCTTTTTTGATTACTTAGAATTTTTTAAAGAGCTCT
TTTAGAGTTAGACATAGATTTTTGTCATATGTCACTTGCACATTCAATAAGATGGAAAAC
[A,C]
CAAGTGAAAAAACACATAAGGAATTGCTAAATTTCACATATTTAGAGTCTGCCTTCTGAA
TTGTTTTTGGAGTCAGAGTTGTTAATACCTGTAATTTTCCGTTAAACATCCTCTGTGCCG
CCAAGAGAATTGGTGATGTAGCATTCCTTTCAAGATCCCAAAAAAGAATGCGAAGGTTTT
GGTGCTGGCCTTCAGCTTTGCAATTATGCAAAGCCAGCCTACTTTGACTGCTGCTTAGGG
ATTCCCCATCTTCTACTTCCTTCCCAGTCCATTTGGTTCCTAGAGGGTGAAATGAATGCT

13560 TTTCAGGCTGTTGACTGTCATATGCAAATGTCATGCTGGCAGTTTTGTTATTTTCCCATG
TGTAAGCAATGACAACATCATAATTGGCTTCTGTCTGATAGCAATTGTAAGAGGAATCCC
AATTTCTGAAATGTTACCCAAAAAAGTGACTTTAATTGACGAAGTATGATGATGTAGAAG
GATAGGCAAGAAATGCAAAAGGTAATTTAGAAAGGTTTCATGGGTAAAATGTGACCTATG
TGATCTAGGGCTATAAAGGATTTCAATAAGCAGAAGCACGAGGTGGGTTGTTGAAGAAAG
[C,A]
ACTAAATGTTTTTGGATAAAGAATATAATAATTTGAGAGTAAAGGGTAGAGGGAGGGTTA
TGTAGGTAAGTAGTTGTAAGATGGGGAAAGATTGGGTAGTATTTAGCATTTATCCTTAAT
GTTGACTTCAGTGTAGTTCTCTTTGTGTGTTTTCTAGTATAAACTGCATACATGAAAGTT
AAGAATCTTGTGTTAAGTCCCATATAGGAAGGAAGTAGATAGGAAAACCAAACTGGAAAA
ATGTATGGAGATGTTGGTGAAATGACAGGAACGAAAGCAGCTTGTCTGAGCTTGATCTCT

14127 AGGAACGAAAGCAGCTTGTCTGAGCTTGATCTCTTCACTTCCTCAGTGGTGGTTCTGAGC
GCTGGTTTGGCTGAACTCCACTTACCAGGGAAAAGGGCATAAAGTAAACAGGGTTTGTGT
GGAAGAAGTGGAGTAGAACAAAGTGGAGAGGATCTCTGTTCATTTAGTGTATCTGACAGT
GTGCTTGTCAAGTCATAAAACACTTGAGGATGGAAATCTGGAAGTCATTGTATACATTTT
CTTCTTTCCCTAACATCTAGTCAGTTACAGTTTCTGCCAGTTCTTTTGCTTTTTCCATGT
[T,G]
TTTGGAGGCTGTTCCTCTTCGCTCCACATGTAGTAAATGCTCTAGTTCATGACCCATGTC
TTATCTGGACTGCCATGTCAGCTTCCTAACTCATCCATTCACAGCACCAGTGACTGTAAA
ACAGCATTAGTGAGGATAAAACAGTGGCTGTCAAACTTTTTTGACTGTGGCCCCCAGTAA

FIGURE 30

AAATACACTTTGTATTGCAACTTATGTATACTTTATATATGTATGAATAATTAAAACAAA
AGGTTGATTCAAGAAAAATCTTTACATTTACCCTGTGCCATGCAATCTTATATCTTGTAT

18789 GGATACTGTTATGTATCATTTCACTTATATTGGCTTCAGCTTGGGATTTCTCTACTGTAA
GTGGTGAGAATTGATCAGATAGTTAAGGAAGGTCCTTAGATAATGCAGTATACTTATTAA
CATACAGACATCAAGAAGCAGAAATATATAGACATCTTCCTTTTTGGTTCTAATAGGGCT
TCGTGGGACACATATGCAACATGCCTATGATTTTTACAAGCCTGATATGCTATCTGAATA
TCCTATAGTAGATGGAAAACTCTCCATACAGTGCTACCTCAGTGCATTAGACCGCTGCTA
[T,C]
TCTGTCTACTGCAAAAAGATCCATGCCCAGTGGCAGAAAGGTAAGTTTTACCCATTTTCC
TTGGTTTTGGTATGAGTTGAGAGCAGTCTAATGTACTAGGTATCTTTGGTAGGCAACTAC
TTTGTGGGCATTCTTCATTTAATATCCTTTTACCATTAATTCCTCATTCACCAAACAACA
TTTTCCCATAGTTTCTGGGAAAGTGTAATTTACTAGAAGAGGTAAACTTTGGAACTGAGG
TGTATCTCTGCAAAAATATTTAGGTCGGTTTACCCCTTGTAAGAAAATCAAAGTGGAGAA

18901 CTTATTAACATACAGACATCAAGAAGCAGAAATATATAGACATCTTCCTTTTTGGTTCTA
ATAGGGCTTCGTGGGACACATATGCAACATGCCTATGATTTTTACAAGCCTGATATGCTA
TCTGAATATCCTATAGTAGATGGAAAACTCTCCATACAGTGCTACCTCAGTGCATTAGAC
CGCTGCTATTCTGTCTACTGCAAAAAGATCCATGCCCAGTGGCAGAAAGGTAAGTTTTAC
CCATTTTCCTTGGTTTTGGTATGAGTTGAGAGCAGTCTAATGTACTAGGTATCTTTGGTA
[A,G]
GCAACTACTTTGTGGGCATTCTTCATTTAATATCCTTTTACCATTAATTCCTCATTCACC
AAACAACATTTTCCCATAGTTTCTGGGAAAGTGTAATTTACTAGAAGAGGTAAACTTTGG
AACTGAGGTGTATCTCTGCAAAAATATTTAGGTCGGTTTACCCCTTGTAAGAAAATCAAA
GTGGAGAAAAGAAGGTAAGTTGAATTTTGTTCATCTTTTGAGAGAGGTATTTTAACAAGG
TTTTGGACTACAGCTGTGATTCAGGGAAAGCTAATGAAAATGAATTACTAAAGTGATCTT

22095 ACAAGAATGTATTGAGAACTGAGTCCAGGCACAGTGGCTCACACCCGTAATCCCAGCAGT
TTGGGAGGCCGAGGCGGGCAGATCACCTGAGGTCAGGAGTTCGAGACCAGTCTGGCTAAC
ATGGCTGAAACCCCATCTCTACTAAAAATACAAAAATTAGCCAGGTGAGGTGGTGCATGC
CTGTAGTCCTAGCTACTTGGGAGGCTGAAGTAGGAGAATCACTTGAATCCAGGAGAGGGA
GGTTGTGGTGAGCCAAGATCACACCACTGTGCTCCAGCCTGGGTGACAGAGCGAGACTCT
[G,A]
TCAAAAAAAAAAAAAAAAAAATGTATTGAGAACTACTCTGGGGAAGTTGATTTAGCAGTCT
TCTCAAGTGAGCACCTGAATCTGTCCCACAGATCATTACAATATTTTAGTCTTCATTACT
TCTTTCAGTAGGTTTTTACTCTCTGCCCTAAAAATCTATCCAAAAAAAAAAAAAAAATTCT
ACCTTATCTGGATAAAGGATAGGACTAAGTTATCTAATTTTTATAGGCTTATGGTCTTGG
CTATATTTAAGGTCACTTTTGTGCTTTCCCTGAGCAGGAAAGAGCAAAAATGTAGAGATA

22257 AGGTGAGGTGGTGCATGCCTGTAGTCCTAGCTACTTGGGAGGCTGAAGTAGGAGAATCAC
TTGAATCCAGGAGAGGGAGGTTGTGGTGAGCCAAGATCACACCACTGTGCTCCAGCCTGG
GTGACAGAGCGAGACTCTGTCAAAAAAAAAAAAAAAAAAATGTATTGAGAACTACTCTGGG
GAAGTTGATTTAGCAGTCTTCTCAAGTGAGCACCTGAATCTGTCCCACAGATCATTACAA
TATTTTAGTCTTCATTACTTCTTTCAGTAGGTTTTTACTCTCTGCCCTAAAAATCTATCC
[-,A]
AAAAAAAAAAAAAATTCTACCTTATCTGGATAAAGGATAGGACTAAGTTATCTAATTTTT

FIGURE 3P

ATAGGCTTATGGTCTTGGCTATATTTAAGGTCACTTTTGTGCTTTCCCTGAGCAGGAAAG
AGCAAAAATGTAGAGATAAACTGATGAAAACTTGACATTACTTTTTAAAATTATACCATG
GGCCAGGTGCAATGGCTCACACCTATAATCCCAACACTTCAGGAGGCTGAGGTGGGAGGA
TTGCTTGAGGCCAGATGTTCAAGGCCAACCTGAGCAACATAGTGAGACCCCATCTCTATA

22582 TCTGGATAAAGGATAGGACTAAGTTATCTAATTTTTATAGGCTTATGGTCTTGGCTATAT
TTAAGGTCACTTTTGTGCTTTCCCTGAGCAGGAAAGAGCAAAAATGTAGAGATAAACTGA
TGAAAACTTGACATTACTTTTTAAAATTATACCATGGGCCAGGTGCAATGGCTCACACCT
ATAATCCCAACACTTCAGGAGGCTGAGGTGGGAGGATTGCTTGAGGCCAGATGTTCAAGG
CCAACCTGAGCAACATAGTGAGACCCCATCTCTATAAAAAATAATAAAAATAAAATAATT
[A,G]
TACCATGGATTAATTGTAGACAAGTTATTTATAGTTTCAAATTATGCCTGTTTCCTAACT
TGTCTAGTGGCAGATACTCAATAATAGATTTCTAGTCTGACATCATAGGAGATTTGTCAA
ATAGGTATCATCTTATCTTTTAACTAATCAGTAGCCAGTAGTTTTAATGAAAATGAAAAG
TTGTTTTGCCTCATTTGGCAACATTTTACTTAGGCTTCTTTTGGACATGATTTTTCAAAA
AAATCTTTTAATGTTGAATTATTCACTATTTTAGGGTCTGCTCTTGATAAAATAACAGCA

25232 ATGTTAAGGGCTCTGTAAAACTTCATACCTCTTTGGCCATTTGTATGCATGATGTTTGGT
TTTTAAACATGGTATAATGAATTGTGTACTTCTGTCAGAAGAAAGCAGAGGTACTAATCT
CCAATTAAAAAATTTTTTAACATGTAAGAATTTTGTACTTTGAACAACAAGATTACAGAA
AGTACCTGTGGTTTTTGGAAAACATTTCTAGCTTGGGGAATGTGACAACATTCCCCAGTG
TGGTAAAATTGGGGTAAAATGTGGTAAAATGTGATACGCACAAACCCTTTGAAAATAGCA
[A,T]
AACAAACATGCCCTTTTTCTAAAATTGATAAATCCTAAAGAGGAAGAAAAGAGCTGGGAC
AATAAAACACTGGCTCTGGAATCTGGAATGTTAAGTCCAGGCCAGCAGTGACAAAAGTTA
TTGTAATGACCTCTGAACAGAGAAACACTGCCATTGAAGAGGCTTCTGGTATAGAAAACA
TGGTACATTCAGGAGCTGTGAATATAGCTCTAGGTGTGCTCCTGAATCAGTTCATGGTAG
ATTATGCTGAACAACAGTGAGATGTTATTGGAGGTGTGGATGAGGGAGTTTGTTGTTGCA

26224 CATGAGCCTTATAGATTGTAAAATAGGGATAGTTGGAACTAATGTACAGAACTAAATTTT
TTAAACTTTATTTGCTGTTAAATTCTGTGAAGTTTCAGTTATCTAAAATAAATATACACA
AATATGAAATATAATGTTTCAGATTGCAAGGTAATATGTAATAGTAGTGTTTGTAAGATA
CTCTTGTCTAATATTAACTAGTAGTATTTTGATTTGTACAGTCATAATTTGTTAAAATGA
CTTCATTTAACATTCACTGATGTAGATTAATAATGTAAGTTCTGATTTAAAGAATGGTGG
[C,G]
AAAATGGTGCATGTAATACTTTTGCAAGTGTTGGGGAGATCGGTATGTTTTGAAAAGAGT
AATTTAACTTTTGGGTGCCAGGAAATGGGTTTTCTCAAAGTCCATTGCCGGCAATGGCA
GGCCTGCAAATACTGGCACAGAGCATTAATCATACACCTTATTAACGGTGAGGTGAATAA
CTTTGAAATAAAGTTTTAGAGAAATGTTTCAGATACTTGAGTATTCTTTTTCACTCTTGA
ACTAACAACTTCGGCAAGAAATCAGCTAATATTCTATTTTTAAATATGGGCATTAATTTC

26695 AGGTGAATAACTTTGAAATAAAGTTTTAGAGAAATGTTTCAGATACTTGAGTATTCTTTT
TCACTCTTGAACTAACAACTTCGGCAAGAAATCAGCTAATATTCTATTTTTAAATATGGG
CATTAATTTCATTTCAGTTCGTTCACTCATTCCATTCATTTATCATTTCACAAACATTTG
AAATCCTAATATAAGCAAGGTGCTCTGTTTAAGGCAGAAATTTGAAAATGTACAAGATAT
ATGGTCTTGTCTTTAAGGAGCTGTTCATCTAGAATGGAGGAATTTACACTGATAATTATT

FIGURE 3Q

[C,T]
CTACACTTGAAACAAAGAAATTAACTCTCAAATTGCGTGGCAAGCATATATAGACTTTGC
TATAAATATTTATGAAATGAGTTACTGTTTTCCTTAAAAAAGCTAAGACTAAGGGCTGGC
AATCAAATAAGAGCAAATTTAGTGGTGAACGTAGAACTGCCCACTACCAGCTAGAGTCTC
CAACCTAAAAGTCCCATGTTGCTAGTGATCCCCAGGGGTTTTATAGAAGGAATCCCTGCA
TTGGCAGTAATTTTGGATTAGATGATCCCTAAGAGCACCATCAAGTCTTAGGATTCTATG

27982 TTGAACGAGGTAAGTGCTTGGGAAAGCATTTTTGTTTTTTTTAGCACAATATGCTGAGAA
ATTTGAAAATAGAAGTAGGAGCTGTCGCTTACTTAATGGTCATTAAATGCAGGTACTACT
TGCTAAGAGCTTTATGTGTGTTATCATATTTATGTTTTTTTTTCTTTTTTTTTTTTTTTG
AGACCGAGTTTCACTCTTGTTGCCCAAGCTGGAGTGCAATGGCACGATCTCGGCTCACTG
CAACCTCTGCCCCCAGGTTCAAGTGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGATT
[A,G]
CAGGCACACACCACCATGC

FIGURE 3R